

SYSTEM AND METHOD FOR PACKING OF SMALL OBJECTS FOR EFFICIENT DELIVERY OVER NETWORKS

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to network data communications methods and systems. More particularly, it relates to client-server data communications and the efficient delivery of objects identified in a search.

Description of the Related Art

The Internet, as it is popularly known, has become an important and useful tool for accessing a wide variety of information. One component of the Internet is the World Wide Web (hereinafter the web). In recent years the web has become an increasing popular vehicle for providing information to virtually anyone with access to the Internet. Many sites (hereinafter, websites) have been established to provide, over the web, information in many different forms, such as text, graphics, video and audio information.

A typical website is hosted on a network server computer that includes application software programs. The server, also known as a web server, is connected to the Internet. By connecting the web server to the Internet, clients that are connected to the Internet can access the website via the web server. Usually, a client is located remotely from the web server, although the client and the server can be at the same location. A web server also can be connected to a private intranet, as opposed to or in addition to the public Internet, in order to make a website privately available to clients within an organization. A conventional

1 client-server communication system used on the web is shown on Fig. 1. The system
2 includes a client communicating to a server over a network connection.

3 The client accesses the website by using a web browser. The web browser is a
4 software program that runs on the client and receives from the server information formatted
5 in a known manner. A very popular format for information sent over the web from a server
6 to a client is the hypertext mark-up language (HTML). HTML is a tag-based script
7 language, in which tags surround the information to be presented. By tagging the
8 information to be sent to a browser, the browser can interpret the tags and handle the
9 presentation of the information sent from the server. It is left to the web browser to
10 determine the specific display formatting of the information, based on the tags included in
11 the HTML file sent from the server. For example, information to be displayed at a client
12 might include text and graphics. That text, and even the graphics can include, by way of the
13 HTML tags, links to other data, where the links are associated with the web page being
14 presented by the browser. The browser communicates with the server using the hypertext
15 transfer protocol (HTTP). Accordingly, the client includes an HTTP software module for
16 communicating over the Internet with the server.

17 The server includes a transaction processor as well as software for delivering an
18 object and accessing an object. The server also has access to one or more data stores for
19 storing objects. The server includes an HTTP module that operates according to the HTTP
20 protocol, as is well known in the art of data communications over an Internet.

21 A user, through the client's browser, can request a web page stored at the server in
22 one of the object stores. The server then locates and returns the web page to the client for
23 display on the browser.

Many types of search engines are in use for locating information among the vast amounts of information accessible on the Internet. A search engine allows a user to search for occurrences of various types of information, such as text, graphics, etc., that are present and accessible through the Internet. A server, upon receiving a search request from a client, creates an HTML page that is returned to the browser. Displayed in the HTML web page are the results of the search, also known as search hits. These search results can include a list of text items describing the search hits. Alternatively, a server can return the search results in the form of graphical images, as opposed to a textual list, where the information to be located is graphical in nature. The returned graphical images typically are miniature renderings of larger graphical images, and are referred to as thumbnail images. The thumbnail image is used as a link to the full size version of the image. The server communicates the thumbnail images to the client browser by first sending to the browser an HTML page that includes links to the various thumbnail images. The browser, upon receiving the HTML page with the links to the thumbnail images then requests those thumbnail images, or more generally objects, individually.

The transmission control protocol (TCP)/internet protocol (IP), commonly referred to as TCP/IP, is a suite of protocols and services that are used to manage network communications and applications over the Internet. The TCP protocol is a transport layer protocol that provides a connection-oriented protocol that usually guarantees delivery across the network. To communicate using TCP/IP it is necessary to first set up a session when an object is to be requested. When negotiating a network session, the phase at the beginning of the communication process is called the setup. At set-up the protocol details, communication rates and error handling approaches are worked out, allowing the connection

1 to proceed correctly and reliably. Once a session is setup, the TCP/IP protocol can be used
2 to transport requests from the client to the server and responses from the server to the client.

3 Once the desired communication between client and server is complete, the client
4 and the server agree to stop communicating and then systematically tear down the session
5 and recover port addresses and other resources used for the session. When requesting a
6 plurality of objects from the server and returning those objects to the client, the client and
7 server often establish sessions for each one of those transactions. Establishing those
8 sessions requires incurring the overhead of setup and tear down for each of those sessions.
9 Accordingly, when a search results in identifying a plurality of thumbnail objects to be
10 requested from a server many communication sessions are set up to transmit the objects
11 individually. Consequently, those communication sessions also must be torn down,
12 resulting in a high degree of overhead processing.

13 As search engines and applications deliver a plurality of objects, such as graphical
14 objects, as opposed to a single page of search result text being returned, the overhead
15 associated with returning those objects becomes more expensive both in terms of time and
16 resources. Accordingly, there is a need to minimize the overhead involved in requesting and
17 returning a plurality of objects, in a data communication system, such as the World Wide
18 Web.

19 SUMMARY OF THE INVENTION

20 Therefore, in light of the above, and for other reasons that will become apparent
21 when the invention is fully described, an object of the invention is to reduce the overhead
22 associated with requesting and transferring a plurality of objects from a server to a client.

1 The above and still further objects, features and advantages of the invention will
2 become apparent upon consideration of the following descriptions and descriptive figures of
3 specific embodiments thereof. While these descriptions go into specific details of the
4 invention, it should be understood that variations may and do exist and would be apparent to
5 those skilled in the art based on the descriptions herein.

6 BRIEF DESCRIPTION OF THE DRAWINGS

7 Fig. 1 depicts a conventional client-server system in which a plurality of objects are
8 requested and returned using a plurality of communication sessions.

9 Fig. 2A illustrates a series of requests and responses between a client and a server
10 using conventional techniques that use a plurality of responses deliver a plurality of
11 requested objects.

12 Fig. 2B illustrates a sequence of interactions between a client and a server in which a
13 plurality of sessions are used to request and return a plurality of objects.

14 Fig. 3 illustrates a client-server system in accordance with certain aspects of the
15 invention in which a plurality of objects are returned in a compressed bundle.

16 Fig. 4 illustrates an exemplary sequence of communications between a client and a
17 server, according certain aspects of the invention, in which a plurality of requests are made
18 for objects and a single compressed object packed in a container is returned to the client.

DETAILED DESCRIPTION

Preferred embodiments according to the present invention are described below with reference to the above drawings, in which like reference numerals designate like components.

In a conventional client server system, as depicted in Fig.1, a client 1 includes a browser 2 and an HTTP module 3 for communicating with a server 4. The server 4 includes an HTTP module 5 for handling the communication sessions between the client and server, and is similar to the client HTTP module 3. The client sends requests for objects by way of the HTTP modules 3 and 5. A transaction processor 6 within the server receives the transaction request and processes it to determine the subject matter of the request. The transaction processor 6 interacts with object access module 7. The object access module receives a request from the transaction processor for a particular object stored on one or more of a plurality of object stores 9a-9c. The object access module 7 retrieves the requested object and provides it to an object delivery module 8 within the server. The object delivery module 8 formats the object and prepares it for delivery over a communication path 10, using, for example, a TCP/IP session, by way of the HTTP modules 5 and 3. The individual object requested by the client is sent from the server to the client by way of the HTTP module 3 and provided to browser 2 for display to a user.

Fig. 2A shows an example of a sequence of interactions between the client and the server. Here, the client opens a session, in this case a TCP session between the client and the server. Once the session is open the client then generates a search request that is provided to the server. The server processes the search request by determining search results and preparing those results for delivery to the client. The server returns the results

1 using the TCP/IP session with the client. The client receives information satisfying the
2 search request, referred to as a list of hits. The client processes the list of hits to determine
3 specific objects, e.g., thumbnails, to request from the server. The client then prepares a
4 series of requests for the objects to be retrieved. In the example illustrated by Fig. 2A, the
5 client requests N objects, i.e., object 1 through object N. For each object, the client prepares
6 a separate request for each object. As shown in Fig. 2A a total of N requests are sent from
7 the client to the server. The server processes, with its transaction processor 6, each of the N
8 requests, obtains the N objects by way of object access module 7, and sends each of the N
9 objects to the client, as shown in Fig. 2A. The client then processes the returned objects,
10 such as by displaying them using the browser 2. Once the client receives the last requested
11 object the client ends the session with the server. In this case, although only a single session
12 is used, the server must generate N responses for each of the N separate objects in response
13 to the N requests from the client. By returning each of the requested objects individually, a
14 large amount of overhead processing is associated with responding to each of the requests.
15 Further, the client must individually process each of the N returned objects.

16 Another example of a conventional sequence of interactions between a client and a
17 server is shown in Fig. 2B. Here again, the client opens a session with the server and
18 generates a search request that is provided to the server. The server processes the search and
19 returns results to the client in the form of a list of hits matching the search parameters. In
20 this case, the client ends the session upon receipt of the list of search hits and sets up and
21 uses individual sessions to request each of the identified objects. As shown in Fig. 2B the
22 client transmits a request for object 1, which is listed in the list of search hits, once the
23 session with the server is open. The server receives the request and through its transaction

processor, object access module, and object delivery module retrieves the requested object 1,
prepares it for delivery to the client and returns it to the client. Upon the client receiving
object 1 the client ends the session.

In the same manner, the client then requests each of the remaining objects that are in
the list of search hits. Alternatively, a user might select only certain ones of the objects
listed as hits in the list returned from the server. In that case, where there are M hits in the
search result list, the client requests N of the objects, where $N < M$. With respect to object 2
the client again opens a session with the server and requests object 2. The server returns
object 2 to the client and then the session is torn down and ends. The sequence of opening
sessions, requesting an object, returning the object and ending the session is repeated until
all N requested objects are returned. As can be appreciated, the technique shown in Fig. 2B
requires a substantial amount of overhead in opening and closing a session for each object
requested. This overhead requires the availability and use of resources at both the client and
the server and takes time in setting up and tearing down the many sessions.

The objects returned by the server can be in virtually any format. While the objects
can be text objects for display on the browser, the objects also can be other types of objects
such as graphical images, e.g., thumbnail images, for delivery and display on the browser.
By displaying thumbnail images to depict the search results, a user is given a visual
indication of the objects that can be requested and retrieved. Upon selecting an object for
retrieval, such as positioning a mouse cursor over the thumbnail image of interest and
selecting it using the mouse, a full size object corresponding to the selected thumbnail is
requested and sent from the server to the client. Still other types of objects can be used to
present the search results such as other graphics objects, for example, GIF, JPEG and BMP

1 files, banners, flash objects and other web icons, as well as other types of objects such as
 2 rich media objects, e.g., audio clips, etc. Essentially, any type of object that can represent
 3 information can be returned for presentation at a client.

4 A client-server system shown in Fig. 3 avoids the need to setup and teardown
 5 individual sessions for each of the requested objects. Using the system of Fig. 3 also can
 6 avoid having to return a plurality of objects individually, thereby minimizing the amount of
 7 overhead processing both at the server and client in handling those objects individually.

8 In Fig. 3 the client preferably includes a plugin module 20 that operates with the
 9 browser 2. The plugin module runs on a processor and includes an unpacking module 21 and
 10 a decompression module 22. The server 4 also includes, in addition to the transaction
 11 processor 5, a request processor 30 configured to process the individual requests received
 12 from the client. An object packing module 31 is configured to operate in conjunction with an
 13 object compression module 32. The object compression module 32 operates to compress
 14 objects retrieved from object stores 9a-9c by way of object access module 8. The object
 15 packing module 31 operates to pack the compressed objects into a single response message
 16 for delivery to the client. Preferably, the object compression module 32 compresses the
 17 objects before they are packed, although alternatively, the packed response message that is
 18 output from the object packing module 31 might include uncompressed objects. In that
 19 case, the object compression module 32 can be configured to compress the packed response
 20 message after the objects are packed into that response message.

21 The response message is passed from the object packing module 31 to the object
 22 delivery module 6. The object delivery module 6 interacts with the HTTP module 7 for
 23 delivering the packed response message to the client over the communication path 10.

1 The client 1 upon receiving the packed response message by way of the HTTP
2 module 3 passes the packed response message to the plugin module 20. The plugin module
3 20 receives the packed response message and unpacks it using the unpacking module 21 to
4 recover the individual objects. If the recovered objects are compressed the decompression
5 module 22 decompresses the unpacked objects. Alternatively, the decompression module 22
6 can be configured to decompress the response message and then the unpacking module can
7 unpack the decompressed response message, if the compression was performed after the
8 objects were packed into the response message in server 4. The browser 2 then displays the
9 unpacked and decompressed objects.

10 Alternatively, the plugin module, by way of the processor, can further process the
11 unpacked requested objects. That is, the processor can place the unpacked objects in an
12 order for handling accordingly to an algorithm within the plugin module. For example, the
13 plugin can arrange the unpacked objects for display in a particular order on the browser 2.
14 Conventionally, the objects are displayed on the browser in the order they are received from
15 the server, because they are not delivered to the browser in a single bundle. Such an order,
16 however, is not necessarily the order in which the objects are intended to be displayed.
17 Delivering the objects in a bundle also enables other processing to be performed on the
18 unpacked objects.

19 Fig. 4 shows an exemplary sequence of messages between a client and server using
20 the client-server system shown in Fig. 3. Here, the client begins by opening a session with
21 the server and generating a search request. The server processes the search request,
22 determines the search results and returns those results to the client. The client then requests
23 one or more of the objects designated in the search results. Assuming, M hits are returned in

1 the search results, for example, the user can request up to M of the objects from the server.
2 Here, the user requests N of the objects, where $N \leq M$. Those N requests can be made
3 individually or they can be bundled into a single request by the plugin module. Where the
4 objects are bundled into a single request, the plugin module includes a client bundling unit
5 or a packing unit similar to the object packing module 31 in the server. The client packing
6 unit packs the requests into a single message and sends a single message from the client to
7 the server to retrieve the N. In that case, the server 4 includes unpacking and decompression
8 modules similar to the unpacking module 21 and decompression module 22 in the client 1.

9 The server, if receiving individual requests from the client, processes each of those
10 requests using the request processor 30. However, rather than individually responding to
11 each of the requests the request processor operates in conjunction with the object packing
12 module 31 and object compression module 32 to compress and pack the requested objects
13 obtained by way of object access module 8 into a single container. The single container,
14 such as a response message unit, is then returned from the server to the client. The client,
15 upon receiving the response message, passes the message to the plugin module for
16 unpacking and decompressing the message. Upon receiving the response message the client
17 ends the session. The objects are then unpacked, decompressed and transferred to the
18 browser for presentation to the user. Although the figures show performing the unpacking
19 and decompression after the session has ended, alternatively, those operations can be
20 performed before the session has ended, however it is not necessary to do so.

21 As can be seen in Fig. 4 a single session is used to request and receive search results
22 and to request and receive a plurality of objects. Alternatively, a plurality of requests can be
23 made to request each of the N objects using a plurality of sessions while using only a single

session to return the single response message containing each of the requested objects to the client.

In the manner described above, the client-server system can operate efficiently and effectively when a plurality of objects is requested from the server for return to the client.

The response message output from the server to the client includes data fields sufficient to delineate each of the objects returned within the response message. The specific layout and format of the data fields used in the response message will depend on the packing techniques employed by the object packing module. Preferably, the response message includes a directory area having information identifying and describing the objects packed into the response message. Preferably, the directory includes a directory entry for each object packed into the response message. A directory entry can include an object name or ID field containing the object's name or ID, a type field (mime) with information specifying the type of object, and a size field indicating the size of the object. Still other information can be included in a response message directory depending on the needs and capabilities of the browser.

For example, the directory can also include, either for each directory entry or for the entire packed message, a presentation order field indicating the order in which the packed objects are to be presented by the browser. The browser queues the requests waiting for the respective responses. When the browser receives a response to a request it processes that response and presents it to a user upon receipt. When the plugin module unpacks objects from a response message and transfers an unpacked object to the browser, the browser receives the object as if the object were returned by itself over the network in response to a queued request. The browser then presents that object once it receives it. By the plugin

1 module transferring an object to the browser in the order specified by an order field in the
2 response message, the plugin module controls the order in which objects are passed to the
3 browser, which thereby controls the order of presentation of the objects by the browser.

4 An example of an object packing module 31 that can be used in server 4 is software
5 program that packs the retrieved objects into a ZIP format using known ZIP packing
6 techniques. Similarly, the object compression module 32 can use well-known compression
7 techniques to compress the objects either before packing or to compress the response
8 message once the objects are packed into the response message. In the case of using a ZIP
9 packing module the objects reside in their respective formats on the object stores 9a-9c and
10 are packed together by placing them into a ZIP object in response to the request processor
11 30 requesting the specific objects requested by client 1. Packing the objects occurs
12 automatically in response to and under the control of the request processor 30.

13 Once the packed response message is sent to the client 1 and upon the plugin module
14 receiving the response message the plugin module 20 automatically passes the zipped object
15 to the unpacking module 21. The unpacking module 21 automatically performs an unzip
16 operation, if the server object packing module 31 packed the objects in a ZIP format, to
17 thereby unpack the objects within the ZIP object. The decompression module 22 then
18 decompresses any objects that have been compressed. These operations are performed
19 automatically without user intervention, unlike conventional uses of ZIP files in which a
20 user manually zips, or places, files into a ZIP file and then manually controls the unzipping
21 of the ZIP file. Moreover, it should be noted that the object packing module 31 under the
22 control of request processor 30 dynamically builds the ZIP object on the server without user

1 interaction. Such a ZIP object is built dynamically using the objects requested by the client
2 1.

3 Alternatively, the object store can store a static packed object having a plurality of
4 objects that have been prepacked and precompressed. Such static packed objects can be
5 used, for example, to respond to a predefined query from the client. In such an embodiment
6 in which the objects are packed prior to receiving a request, the client plugin module 20
7 detects that the response message contains objects packed in a predetermined format and
8 automatically unpacks the message using the appropriate unpacking algorithms. An
9 example of a use of static packed objects arises when a web page having a plurality of static
10 objects, such as thumbnail images, is predefined for searches that are preformed frequently.
11 The thumbnail images for the web page are known prior to receiving a search request.
12 Accordingly, those thumbnail images are statically compressed and packed into a single
13 object, ready for delivery when a request is received by the server. By the server storing
14 static packed objects, browsing such static thumbnail web pages is greatly accelerated. This
15 is because the process of the server gathering, compressing and packing all of the requested
16 objects is performed only once when the static web page is first defined, and is not
17 performed at the time the request is received.

18 It will be understood that the request processor 30, object packing module 31, object
19 compression module 32, as well as the transaction processor 5, object delivery module 6 and
20 object access module 8 can be implemented as software programs. Such software is
21 embodied on a computer-readable medium of expression, such as magnetic media, optical
22 disks, semiconductor memories, etc. The software instructions within these modules, when
23 executed by a computer, cause the computer to perform the functions described above. It

1 will also be understood that the functions described above can be performed using
 2 appropriately configured hardware or combinations of hardware and software. Likewise, the
 3 plugin module 20, including the unpacking and decompression modules, preferably are
 4 implemented as software with computer-readable instructions recorded on a computer
 5 readable medium such as magnetic media, optical disks, semiconductor memories, etc.

6 Having described preferred embodiments of systems and methods for packing
 7 objects for efficient delivery over networks, it is believed that other modifications, variations
 8 and changes will be suggested to those skilled in the art in view of the teachings set forth
 9 herein. It is therefore to be understood that all such variations, modifications and changes
 10 are believed to fall within the scope of the present invention as defined by the appended
 11 claims. Although specific terms are employed herein, they are used in their ordinary and
 12 accustomed manner only, unless expressly defined differently herein, and not for purposes
 13 of limitation.